DIVERSITY OF ARCHAEA FROM THREE FORESTED ECOSYSTEMS IN GSMNP

By

Philip Jon Drummond

A Thesis
Submitted to the
Faculty of the Graduate School
of
Western Carolina University
in Partial Fulfillment of
the Requirements for the Degree
of
Master of Science

Committee:

Sean O'Connell _____ Director

Karla M. Weldon _____

Sabine Rundt _____

_____ Dean of the Graduate School

Date: July 31, 2006

Summer 2006
Western Carolina University
Cullowhee, North Carolina

DIVERSITY OF ARCHAEA FROM THREE FORESTED ECOSYSTEMS IN GREAT SMOKY MOUNTAINS NATIONAL PARK

A thesis submitted to the faculty of the Graduate School of Western Carolina University in partial fulfillment of the requirements for the degree of Master of Science.

By

Philip Jon Drummond

Director: Seán O'Connell, Ph.D.
Assistant Professor of Biology,
Department of Biology

July 2006

HUNTER LIBRARY
WESTERN CAROLINA UNIVERSITY

Table of Contents

Page

List of Tables iv

List of Figures v

Abstract vi

Introduction 1

Methods 10

Site Location 10

Sample Collection 10

DNA Extractions 10

Polymerase Chain Reaction (PCR) for Molecular Cloning 11

PCR Product Cleaning 11

Molecular Cloning 12

Whole Cell PCR 12

Restriction Fragment Length Polymorphism (RFLP) 13

PCR for Denaturing Gradient Gel Electrophoresis (DGGE) 13

DGGE 14

Principal Components Analysis 15

Sequencing 15

Phylogenetic Analysis 15

Results 17

PCR 17

Molecular Cloning 17

RFLP 17

DGGE 18

Principal Components Analysis (PCA) of DGGE Banding Patterns 18

DNA Sequences 18

Phylogenetic Analysis 19

Discussion 41

PCR 41

Molecular Cloning 42

Variation in Sizes of Clone Inserts 42

Sequence Analysis 43

Phylogenetic Analysis 44

PCA Analysis 45

Ecological Role of Archaea 46

Conclusions 49

References 51

Appendix 57

List of Tables

Table Page

1. Sample site descriptions 21
2. Primer sequences and targets 22
3. 16S rDNA sequence matches from Ribosomal Database Project II.. 23
4. 16S rDNA sequence matches from GenBank 24

List of Figures

Figure		Page
1.	PCR optimization of all samples	25
2a.	PCR using primer set 46f and 1492r	26
2b.	PCR using primer set 21fa and 1492r	27
3.	Whole cell PCR gel demonstrating insert size differences	28
4a.	RFLP of clones 15 - 40	29
4b.	RFLP of clones 41 - 113	30
4c.	RFLP of clones 116 - 218	31
4d.	RFLP of clones 222 - 318	32
4e.	RFLP of clones 358 - 407	33
4f.	RFLP of clone 410 and clones with various sized inserts	34
5.	DGGE gel of all samples	35
6a.	PCA of DGGE banding patterns for all sites	36
6b.	PCA of DGGE banding patterns for Albright Grove and Cataloochee	37
6c.	PCA of DGGE banding patterns for Albright Grove and Purchase Knob	38
6d.	PCA of DGGE banding patterns for Cataloochee and Purchase Knob	39
7.	Phylogenetic tree for all GSMNP archaeal sequences	40

Abstract

DIVERSITY OF ARCHAEA FROM THREE FORESTED ECOSYSTEMS IN GREAT SMOKY MOUNTAINS NATIONAL PARK

Philip Jon Drummond, M.S.

Western Carolina University, July 2006

Director: Dr. Seán O'Connell

Prokaryotes are vital to the survival of all life on Earth since they control the cycles of many elements including carbon, nitrogen, sulfur, etc. The study of the *Archaea* has resulted in numerous novel metabolic discoveries, most from extreme environments; however, little is known about archaea and their roles in temperate ecosystems. DNA was extracted directly from soil from three forested ecosystems in Great Smoky Mountains National Park and was used to characterize community structure using molecular techniques including PCR followed by molecular cloning and restriction fragment length polymorphism (RFLP) analysis, denaturing gradient gel electrophoresis (DGGE) and DNA sequencing. Seventeen archaea were sequenced, including species aligned to the phylum *Crenarchaea*, which so far contains only one organism cultured from a non-extreme environment. Overlap was seen between clones sampled from multiple sites and from DGGE banding patterns, indicating that some archaeal species are widespread. The extent of archaeal diversity is unknown and is

thought to be dwarfed by bacteria; however, our understanding of archaea is limited due to their resistance to being cultivated. Obtaining a baseline of diversity in this group should ultimately help yield isolated species for further study of their unique metabolisms and biochemical properties.

Introduction

Prokaryotes are one of the most vital groups of life on this planet. Their very diverse range of metabolisms allow them to provide most of the nutrients needed for all other forms of life (Madigan et al. 2003). Without these microbes, all other forms of life would not be able to survive. Their effectiveness at providing the nutrients is mainly based on the large numbers and types of microbes that exist. *Bacteria* and *Archaea* have been found in virtually every environment including soil, river water, lake sediment, marine systems, animal guts and plant and animal systems, etc… They are also found in environments where very few to no other organisms can survive, e.g. deep in the oceans (Teira et al., 2004), hot springs (Stahl et al., 1985), acid mine drainage sites (Bond et al., 2000) and even nuclear waste (Fredrickson et al., 2004). They are so numerous that they comprise about one half of the Earth's biomass, the other half consists of plants. Animals provide an insignificant amount (Whitman et al. 1998). Microorganisms, including prokaryotes, mediate such processes as nitrogen fixation (performed by bacteria and archaea), decomposition in soil and water environments (performed by bacteria, fungi and arthropods) and photosynthesis (performed by bacteria, cyanobacteria and algae) (Madigan et al. 2003).

Arranging organisms into a classification scheme has been in practice for centuries. Originally, Plantae and Animalia were described as the only two divisions of life. This changed when Haeckel, in 1866, realized that a third division should be added to the classification (Woese, 1994). The third division would encompass single-celled eukaryotic organisms, or Protista. A fourth group of organisms were later added and named the Mychota (Copeland, 1938), which were renamed Monera. In 1969, Fungi were added as the fifth kingdom (Whittaker, 1969). Around the same time as Copeland's addition Chatton, Stainer and van Niel suggested that another classification could be used, prokaryotes and eukaryotes (Stainer and Van Niel, 1941). These classification systems were based solely on morphological differences among the individuals within the kingdoms.

It was not until the late 1970's when the prokaryote/eukaryote dichotomy was explored again by the work of Carl Woese and George Fox (Woese and Fox 1977). By digesting the small ribosomal subunits (SSU) of organisms that were representative of eukaryotes, bacteria and methanogenic bacteria with T1 Rnase and comparing the resulting fingerprints, Woese and Fox discovered that the methogenic "bacteria" appeared to be "no more related to typical bacteria than they are to eukaryotes." They proposed to separate these methanogens in a third "urkingdom" named *Archaebacteria* and the bacteria were called *Eubacteria* (Woese and Fox 1977).

With the advent of new and powerful molecular tools came the ability to gain information about organisms without the need to culture them first and this had a very strong influence on the understanding of microbial diversity and phylogeny. These techniques were cited by Woese et al. in 1990 since the growing amount of biochemical, genomic and phylogenetic data supported that the *Archaebacteria* should be considered a separate taxonomic division. The authors proposed that all three divisions should be considered as equal in stature. The new divisions, or domains, are the current paradigm in biology with the three domains named *Archaea*, *Bacteria* and *Eukarya* (Woese et al. 1990).

All of the advances in biochemical, genomic and phylogenetic analyses have further supported the three domain approach to taxonomy. *Archaea* contain general characteristics that are shared with either or both *Bacteria* and *Eukarya* as well as unique characteristics. Features that are shared between *Bacteria* and *Archaea* include small cell size (<10μm); a small, non-membrane bound, circular, single chromosome; division by binary fission; their ribosomes have a combined size of 70S; there is a lack of membrane-bound organelles and some have been shown to fix N_2. *Archaea* share common characteristics with *Eukaryotes* as well, such as the presence of histone proteins and the ineffectiveness of anti-bacterial agents. The characteristics that separate *Archaea* from both *Bacteria* and *Eukarya* include containing cell membrane lipids linked with ether bonds instead of ester bonds, *Archaea* are the only confirmed organisms that can produce

methane, some organisms contain a phospholipid monolayer cell membrane and *Archaea* cannot perform photosynthesis (Madigan, et al. 2003).

When *Archaea* were first described, they had the reputation of being obligate extremophiles, organisms which can grow optimally under one or more chemical or physical extreme (Madigan et al. 2003). This was investigated further and they were found in environments such as those that were anaerobic, thermal (45°C or higher), cold (15°C or lower), very acidic (pH of 2 or lower) or had a very high concentration of salt (15% or greater). This tenet was accepted until *Archaea* were found in open ocean environments (DeLong, 1992). Since then, *Archaea* have been found in numerous non-extreme environments (Simon, et al. 1999; Jurgens, et al. 1999), leading researchers to show that *Archaea* may be just as ubiquitous, if not more so than, *Bacteria*. These results have led to the discovery of novel and diverse biochemical pathways such as autotrophism using rhodopsin (Béjà, et al. 2000), or chemolithoautotrophy by oxidizing ammonia (Könneke et al. 2005).

The domain *Archaea* has been divided into four phyla; *Crenarchaea*, *Euryarchaea*, *Korarchaea* and *Nanoarchaea*. *Crenarchaea* contains all known "non-extreme" *Archaea* as well as some sulfur-dependent hyperthermophiles (Madigan et al. 2003). There has only been one reported non-extreme *Crenarchaea* that has been brought into pure culture (Könneke, et al. 2005). This organism oxidizes ammonia and was isolated from marine sediments. *Euryarchaea* contains all of the methanogens, the halophiles and many of the

thermophiles. The majority of the cultured *Archaea* come from this phylum. The *Korarchaea* contain a small number of organisms, which have all been hyperthermophilic. The most recently discovered phylum, *Nanoarchaea*, contains only one known member (*Nanoarchaeum equitans*) which has only been cultured when grown with *Ignicoccus sp. KIN4/I. N. equitans* is the smallest known living cell with the smallest known genome of 480kb (Huber et al. 2002). All of the studies done thus far have failed to yield an archaeon that is pathogenic to humans (Reeve, 1999).

The introduction of molecular techniques for studying microbial diversity, e.g., polymerase chain reaction (PCR), restriction fragment length polymorphisms (RFLP) and DNA sequencing have allowed researchers to study microorganisms from natural environments in a more effective way. Because of the awesome diversity of prokaryotes and the multitude of niches they inhabit, techniques that involve growing organisms on microbiological media has been a very inefficient way to assess the diversity, since only around one percent of environmental organisms can be cultured (Torsvik et al., 1990). This low number is due to the fastidious nature of the organisms and our inability to precisely mimic their natural environment, in particular, the non-extreme archaea. This is where molecular tools play their role. With these tools, DNA can be directly extracted from the environment, analyzed and ultimately assigned taxonomic meaning by sequencing it.

When attempting to identify novel organisms by comparing the DNA

sequence to a bank of known sequences a standard gene region must be used (Woese 1987). It is much easier to sequence a small region than to sequence a large one, or even the entire genome. As a result, the SSU ribosomal subunit is used. It is an effective sequence to use for many reasons: it is present in all living organisms regardless of the domain, it is very easily extracted from the organism, it is an ancient molecule that has remained conserved in its overall structure and there are regions that have remained strictly conserved among all three domains. However, there are also variable regions that are shared among lower taxonomic levels all the way to sequence variations among individual species. These characteristics serve to identify species and also to evaluate evolutionary relationships.

To obtain the rRNA or rDNA sequence, a series of techniques are used. These studies typically begin with the collection of the environmental samples (Madsen 1998). The DNA is extracted from the samples and the desired sequences are amplified using PCR with primers that are designed to target the ribosomal RNA gene. With the amplified DNA obtained, several tracks can be followed. Techniques designed to separate species within the communities, e.g. denaturing gradient gel electrophoresis (DGGE), can be employed and the diversity can be assessed. The bands from the resulting DGGE gel can be excised and sequenced to determine species composition. PCR products might also be ligated into a cloning vector and then transformed into *Escherichia coli* cells and the cells cultured (Kobs 1997). The inserts are then removed from the

vectors and sequenced. The sequence data can be used to design new primers or hybridization probes to more efficiently identify specific organisms (Madsen 1998). When the sequences have been obtained, they can be analyzed against a variety of sequence databases, e.g. GenBank or Ribosomal Database Project II. The sequences can also be used in conjunction with sequences that are in the database to hypothesize phylogenetic relationships. The subsequent evolutionary tree(s) can be used to evaluate the relatedness to other groups of organisms. Knowing what organisms the sequences are related to can also assist in determining the possible conditions needed to culture the organism (Könneke and others 2005).

Although these molecular techniques are very effective, they are not perfect. They all have their own limitations and biases. Nucleic acid extraction has a limitation in that the quantitative recovery of the nucleic acids cannot be easily assessed. Since it is nearly impossible to know the total amount of nucleic acids present in a sample, prior to extraction, the efficiency of the extraction is difficult to determine (Miller and others 1999). Along with this, microbial spores will be more difficult to lyse than vegetative cells. Gram-positive cells have also shown to be more resistant to cell lysis than gram-negative cells, due to thicker cell walls. To address the first limitation, cell counts can be performed via microscopy before and after the extraction. Performing various techniques to extract the DNA could allow for extraction of microbial spores and gram positive

organisms, but if too many or too harsh lysing steps are employed, the nucleic acids may be damaged (Head and others 1998).

PCR amplification has a bias with selectivity where small differences in the universally conserved regions of the rRNA may cause selective amplification of certain sequences (Head and others 1998). Many prokaryotes are known to contain multiple and/or different rRNA sequences, so the assumption that clone libraries represent *in situ* population densities may not be accurate. In essence there could be a bias towards the organisms that had a greater number of rRNA sequences. As a result of containing more hydrogen bonds between the DNA double strands, there also may exist a discrimination against high %G+C sequences due to a lower efficiency of strand denaturation during PCR. DNA polymerase also naturally makes errors (Qiu and others 2001). The average error rate for a polymerase that is not considered high fidelity is $1x10^{-4} – 1x10^{-5}$ errors $base^{-1}$ so mutations are a likely possibility with amplification. Another issue with PCR involves the formation of chimeras and heteroduplexes. Chimeras occur when a secondary structure forms in the target sequence, which halts the elongation. If this happens several times, then these partially elongated sequences can bind together creating a fragment of DNA that is comprised of the strands of multiple species. A heteroduplex is formed when two single-strands from different organisms anneal together forming a double-strand where each strand originated from different organisms. Heteroduplexes have been formed between sequences with a similarity as low as 76%.

Great Smoky Mountains National Park (GSMNP) was an ideal location to study the diversity of *Archaea* since it is one of the most biodiverse locations on Earth. The park was founded in 1934 and due to the level of biodiversity, GSMNP was declared an International Biosphere Reserve in 1976 (National Park Service 2006). With this recognition, extensive studies to investigate the diversity (other than plants) were not begun until 1998 when the All Taxa Biodiversity Inventory (ATBI) was initiated (Sharkey 2001). The main objective of the ATBI is to catalogue all of the species that reside in the park. Currently the inventory has found over 12,000 species including 1,400 flowering plant species, 4,000 species of non-flowering plants, 200 species of birds, 66 mammal species, 50 native fish species 39 species of reptiles and 43 species of amphibians. As of October 2005, 3,572 species new to the park and 565 species new to science were catalogued as a result of the study. From these new species, 59 bacterial species were new to the park and 97 bacterial species were new to science and only six species of archaea were new to science and one was new to the park, so the study of *Archaea* in the park was virtually untouched (Discover Life in America 2005). All of the prokaryotic records have been contributed by researchers at Western Carolina University. The purpose of this study was to examine archaeal diversity from soil of three forested ecosystems in GSMNP. It was hypothesized that the archaeal communities would be different as a result of the land history of the sites.

Methods

Site location

Three forested ecosystems were selected for study, as classified by differences in their elevation, vegetation, chemistry and history (Table 1). Albright Grove is located in the northeast section of GSMNP. Purchase Knob and Cataloochee are located in Haywood County, North Carolina in the eastern section of the park. All sites were designated as long-term study plots for scientific use by the ATBI.

Sample collection

Soil samples were collected in February 2005 using aseptic technique by removing the leaf litter and any roots with rinsed and flame sterilized tools (small shovel and garden trowel). The soil was homogenized in the upper 12cm of the ground and an aliquot transferred to a sterile 50mL centrifuge tube, where it was immediately placed on ice. Three replicate samples were taken at each site, from around the base of medium Eastern Hemlock (*Tsuga canadensis*) trees.

DNA extractions

DNA was extracted from all nine samples (three from each site) using Mo Bio PowerSoil DNA Extraction kit (Mo Bio, Inc., Solana Beach, CA). The alternative lysis protocol recommended by the kit was used which involved heating the samples at 70°C for five minutes after adding solution C1 (sodium dodecyl sulfate), then mixing well and heating again to 70°C for five minutes

before continuing with the normal protocol as directed. Extracted DNA was screened for quality and quantity using agarose gel electrophoresis and stored at -20°C.

PCR for molecular cloning

Approximately 1,500 base pair fragments of the 16S rDNA genes from the mixed archaeal species were amplified using the universal archaeal primer 46F (Kaplan and others 2001) and the universal primer 1492R (Lane 1991; Table 2). The universal archaeal primer 21Fa (DeLong 1992) was later used along with 1492R as the sole primer set used. PCR reaction mixtures (50μL total) contained 0.05% IgePal (Sigma-Aldrich, Inc., St. Louis, MO), nuclease free water, 1.0X PCR buffer with 2.0 mM Mg^{2+} and 2.5U Taq DNA polymerase (Eppendorf, Inc., Westbury, NY), 0.25μM of each primer (Operon Technologies, Huntsville, AL), 0.25mM of each nucleotide (Eppendorf) and 0.5μL of DNA template. Thermal cycler (Mastercycler Personal, Eppendorf) conditions were 5 minutes initial denaturation at 94°C followed by 30 cycles of 1 minute denaturation at 94°C, 1 minute annealing at 50°C and 2 minute extension at 72°C with a final extension of 10 minutes at 72°C.

PCR product cleaning

PCR products were cleaned using Montage PCR Centrifugal Devices (Millipore Corporation, Bedford, MA) for any product to be used in molecular cloning, DGGE or sequencing reactions.

Molecular cloning

PCR products were cloned into *Escherichia coli* using the pGEM-T Easy Vector System (Promega Corporation, Madison, WI). Clones were stored in LB broth containing 15% glycerol at –70°C. These clones were used for whole cell PCR and RFLP analysis.

Whole cell PCR

PCR was performed using the procedure from O'Connell and others (2003) and consisted of the pre-PCR reaction mixtures (25μL total) containing one half the volume of the 1.0X buffer (Eppendorf) and nuclease free water. The target clone was selected using a toothpick and deposited into the PCR tube with the buffer mixture. The tubes were then placed in the thermocycler where they were heated to 99°C for 15 minutes (cell lysis step). Post heat lysis reaction mixtures (24.5μL total) were prepared consisting of 0.05% IgePal (Sigma-Aldrich), nuclease free water, the remaining 1.0X PCR buffer with 2.0 mM Mg^{2+} and 2.5U Taq DNA polymerase (Eppendorf), 0.25μM of each primer (Operon Technologies) and 0.25mM of each nucleotide (Eppendorf). Thermal cycler conditions were a five minute hot start step at 80°C (where the post heat lysis mixture was added to each PCR tube), 4 min initial denaturation at 94°C followed by 30 cycles of 1 min denaturation at 94°C, 1 min annealing at 55°C and 1 min extension at 72°C with a final extension of 4 min at 72°C.

RFLP

The products from whole cell PCR were digested using 2mg/mL Bovine Serum Albumin, 0.2X Buffer C and 1U/μL Rsa I and Hae III restriction enzymes (Promega Corporation, Madison, WI) (Knittel, and others 2005), following the protocols of R.M. Lehman (unpublished). A total volume of 20μL was used and digest conditions included three hours at 37°C and 15 minutes at 65°C. Four percent agarose gels were used to analyze the digests using run conditions of 1 min at 210 volts repeated three times with a 10 second break between runs and then a 180 min run at 68 volts. Unique banding patterns were selected for sequencing.

PCR for DGGE

Because clone numbers were lower than expected, DGGE was employed to compare archaea from the three sites. Approximately 550 base pair fragments of the 16S rDNA genes from the mixed archaeal species were amplified using the universal archaeal primers 344F with GC clamp and 915R (Casamayor, and others 2002; Table 2). PCR reaction mixtures (50μL total) were as above. Thermal cycler conditions were 5 minutes initial denaturation at 94°C followed by 35 cycles of 1 minute denaturation at 94°C, 1 minute annealing at 53.5°C, 1 minute and 50 second extension at 72°C with a final extension of 7 minutes at 72°C (A-L Reysenbach, personal communication). The products that were not at an adequate quantity for DGGE were concentrated using a Montage cleanup kit spin filter.

DGGE

DGGE methods were adapted from Muyzer and Smalla (1998) and consisted of a polyacrylamide gel impregnated with a gradient of 20% to 60% urea/formamide to which 20μL of community PCR products were added. A Bio-Rad DCode Universal Mutation Detection system (Bio-Rad Laboratories, Hercules, CA) was used to electrophorese samples at 65V for 15 hours at 60°C. Gels were stained with ethidium bromide for thirty minutes, destained for ten minutes and photographed using UV illumination with an EDAS 290 gel imaging system (Eastman Kodak Company, Rochester, NY). To simplify handling of DGGE gels, a UV transparent sheet, the Gel Handler (Sigma-Aldrich), was used to carry and position the gel on the transilluminator. Band locations correspond to unique species, with each sequence becoming immobilized at its mimicked melting temperature in the urea/formamide gradient (higher G + C DNA migrates lower in the gel due to higher energy needed to break three hydrogen bonds versus two hydrogen bonds for A + T pairings). Bands in the same vertical position hypothetically represent the same species, while those that are staggered likely represent different species. Unique bands were selected for a second PCR reaction followed by sequencing by cutting out the bands from the gel using a sterile razor blade and suspending the gel with the band in 10% TE shaken at 150 rpm overnight at 37°C.

Principal components analysis

To compare between the nine samples, a matrix was constructed from the DGGE gel where the rows represented the sample (the DGGE lane) and each column represented a different species. The cells contained a binary set of data where if the species was present in the lane, a 1 was used and if the species was not present, a 0 was used. This data was entered into a principal components analysis using Systat 6.0 (SPSS, Inc., Chicago, IL).

Sequencing

Approx. 550bp of the original 1500bp clone inserts and excised DGGE bands were sequenced using BigDye Terminator Cycle Sequencing Ready Reaction Kit (Version 3.1) and a Model 3130 Automated DNA sequencer (Applied Biosystems, Foster City, CA). The PCR reaction (10µL total) contained 3.2µM of the 915R primer, nuclease free water, 2µL BigDye 3.1 Ready Reaction mixture (Applied Biosystems) and 10 to 20ng DNA template. Sequence comparisons were conducted using the GenBank (Altschul and others 1997) and RDPII databases (Cole and others 2003).

Phylogenetic analysis

16S rDNA sequences were aligned with ClustalX; verified manually; and exported to PAUP* 4.0 (Altivec; D. L. Swofford, Sinauer Associates, Sunderland, Mass.). The phylogenetic tree was generated by using PAUP*. Maximum parsimony and maximum-likelihood analyses were performed. The resulting tree

was evaluated by bootstrap analyses based on 100 resamplings (Simon and others 2005). RFLP pattern Q altered the tree substantially, so it was omitted.

Results

PCR

All nine samples showed amplification of genomic DNA, indicating the presence of archaea in all sites (Figure 1). Original amplification of the samples using primer set 46f and 1492r resulted in successful amplification of the soil samples, but did not amplify the positive control, the archaeon *Halobacterium salinarum* (Figure 2a). Using the forward primer 21fa with 1492r alleviated this issue (Figure 2b).

Molecular cloning

Molecular cloning produced a combined 217 white colonies from the three sites. After a second screening using double the volume of X-gal (20µl of 50mg/ml), 92 clones were found to be white a second time. Whole cell PCR revealed that 41 of the 92 clones produced a PCR product that was the target size, while the remaining amplified fragments were either half the target size, larger than the target size (~2,000bp) or the fragment did not amplify at all (Figure 3).

RFLP

RFLP analysis yielded 19 distinct patterns (Figure 4a-f). Patterns B, C, D and L occurred across multiple sites and one occurred in all sites (B). Patterns A, E, F, G, H, I, J, K, M, N, O, P, Q, R and S were unique to only one site. Albright Grove had six unique patterns (A, E, F, G, R and S), Cataloochee had four

unique patterns (H, I, J and K), while Purchase Knob had five unique patterns (M, N, O, P and Q).

DGGE

DGGE revealed 12 species among the nine samples (Figure 5). Eight species were shared among more than one site and four were unique to other sites. One species was of particular interest because it occurred in all samples, including some from previous work in GSMNP (DGGE band #1). Albright Grove had a range of bands from four to nine with an average of 5.67, Cataloochee had a range of bands from four to six with an average of 5 and Purchase Knob had a range of bands from two to seven with an average of 4.67.

PCA of DGGE banding patterns

Principal components analysis for all three sites based on the DGGE banding patterns showed a weak separation between the Albright Grove samples and the Cataloochee and Purchase Knob samples when analyzed together (Figure 6a). Site versus site comparisons yielded a stronger trend with Albright Grove showing a distinct community from either Cataloochee (Figure 6b) or Purchase Knob (Figure 6c). No discernable pattern was observed between Cataloochee and Purchase Knob (Figure 6d).

DNA sequences

All of the sequences aligned to the phylum *Crenarchaea*. They had similarity values that ranged from 54% to 96.9% when compared to the RDPII database (Table 3) and from 95% to 99% when compared to the GenBank

database (Table 4). Only two of the sequences matched with a similarity value over 90% when comparing to RDPII. The environments that these sequences were most closely related to were associated with soil.

Phylogenetic analysis

Figure 7 shows a phylogenetic tree that includes all sequences from this study and previous GSMNP studies. Along with the GSMNP sequences, other sequences were used as a comparison. They included a cultured thermophile, *Thermococcus celer* (Lepage and others 2004); a cultured halophile, *Halobacterium salinarum* (Bomberg and others, unpublished data); *Euryarchaea* and studies of cultivable non-extreme *Crenarchaea* as well as crenarchaeal clones from forest soils; FRD 38 (Oline and others, unpublished data) and FFSB1 (Jurgens and others 1997); and farm soils, TREC16 (Simon and others 2005). Four distinct *Euryarchaea* clades can be seen, three of which contain only sequences from previous GSMNP work. Nine *Crenarchaea* clades are observable, including the largest clade, clade 8, that includes two of the DGGE bands (found across multiple sites), a single Albright Grove clone and clones from two other GSMNP sites from previous work examining high organic content soil (Alum Bluffs (AB), Beech Flats Prong (BFP) and Purchase Knob (PuK)). Clade 1 contains only one clone from a previous GSMNP study (AB 16). This clone also has the largest branch length value in the tree. Clade 2 includes sequences from two sites and a marine sponge symbiont (*Cenarchaeum symbiosum*). Clade 3 is comparised of clones from four sites, including all three

sites from this project as well as one clone from site AB. Clade 9 is formed by two Albright Grove clones, as well as a clone from farm soil. Clades 6 and 7 contain clones from three sites including two from this study. Clade 5 is the only clade specific to a single site. DGGE 2 and 3 form a distinct clade (4) and include sequences from Albright Grove and Cataloochee.

Table 1. Site descriptions for the three forested sites in Great Smoky Mountains National Park sampled in this study (Mike Jenkins, unpublished data; Sharkey 2001).

ATBI Plot	Albright Grove	Cataloochee	Purchase Knob
Vegetation	Montane Cove	Mesic Oak	Northern Hardwood
Elevation (m)	1034	1382	1529
Soil pH*	4.3	4.3	4.8
Phosphorous (ppm)	18.7	13.3	12
Potassium (ppm)	93.3	81.7	85.7
Calcium (ppm)	224.8	222.8	274.3
Magnesium (ppm)	35.3	35.2	42.7
Organic Matter (%)	3.9	3.8	3.5
Watershed	Indian Camp Creek	Cataloochee Creek	Cove Creek
Geology	Thunderhead Sandstone	Thunderhead Sandstone	Biotite, Augen, Gneiss
History	Undisturbed	Chestnut Blight	Logged

*data collected at time of sampling.

Table 2. Designations and targets of 16S rDNA amplification primers used.

Name	Sequence	Reference
21Fa	5'-TTCCGGTTGATCCYGCCGGA-3'	DeLong 1992
46F	5'-GCYTAACACATGCAAGTCGA-3'	Kaplan and others 2001
344F	5'-ACGGGGCGCAGCAGGCGCGA-3'	Casamayor and others 2002
915R	5'-GTGCTCCCCCGCCAATTCCT-3'	"
1100R	5'-GGGTTGCGCTCGTTG-3'	Lane 1991
1492R	5'-GGTTACCTTGTTACGACTT-3'	"
GC clamp	5'-CGCCCGCCGCGCCCCGCGCCCGT CCCGCCGCCCCCGCCCC-3'	Casamayor and others 2002

Table 3. Ribosomal Database Project II (RDPII) output for archaeal soil clones from Albright Grove, Cataloochee and Purchase Knob.

Clone	Name	Similarity	Environment	Reference
L	FFSA2	80.7	Boreal Forest Soil	Jurgens and others 1999
B	FFSB6	83.4	"	Jurgens and others 1997
A	FFSB2	96.7	"	"
C	FFSB3	88.0	"	"
D	U62812	54.4	Soil	Bintrin and others 1997
E	U62818	77.9	"	"
F	FFSB5	79.3	Boreal Forest Soil	Jurgens and others 1997
G	U62818	79.9	Soil	Bintrin and others 1997
I	Y08985	84.6	Boreal Forest Soil	Jurgens and others 1999
N	FFSC1	79.6	"	"
O	FFSB2	87.5	"	Jurgens and others 1997
P	Y08985	81.0	"	Jurgens and others 1999
Q	U62812	54.8	Soil	Bintrin and others 1997
DGGE1	FFSB2	88.9	Boreal Forest Soil	Jurgens et al. 1997
DGGE2	FFSB11	72.7	"	"
DGGE3	FFSB2	71.8	"	"
DGGE4	FFSB1	70.6	"	"

Table 4. GenBank output for archaeal soil clones from Albright Grove, Cataloochee and Purchase Knob.

Pattern	Name	Environment	Similarity (%)	Reference
L	EV221H2111601H177	Subsurface water, South Africa	98	Gihring and others*
B	"	"	"	"
A	FRD38	Coniferous forest and alpine tundra soil	99	Oline and others*
C	FFSB3	Boreal Forest Soil, Finland	97	Jurgens and others 1997
D	SAGMA-D	South African gold mine water	99	Takai and others 2001
E	FRD25B	Coniferous forest and alpine tundra soil	97	Oline and others*
F	FFSB5	Boreal forest soil, Finland	96	Jurgens and others 1997
G	D_C01	Tropical Estuarine Sediments	99	Piza and others*
I	EV221H2111601H177	Subsurface Water, South Africa	97	Gihring and others*
N	OdenE-150iia	Soil Associated with deglaciation	96	Nicol and others*
O	GFS10-9500ii	Receding Glacier Soil	99	Nicol and others 2005
P	EV221H2111601H177	Subsurface Water, South Africa	97	Gihring and others*
Q	SAGMA-D	South African gold mine water	96	Takai and others 2001
DGGE1	FRD38	Coniferous forest and alpine tundra soil	99	Oline and others*
DGGE2	FFSC1	Boreal Forest Soil, Finland	95	Jurgens and others 1999
DGGE3	NRP-M	Rice Patty Soil	96	Sakai and others*
DGGE4	"	"	"	"

*denotes unpublished data.

Figure 1. Agarose gel of archaeal PCR products after PCR optimization. Lane 1 contains the DNA markers labeled in base pairs (bp). Lane 2 contains the negative control, lane 3 contains the positive control, lanes 4, 5 and 6 contain Albright Grove samples 1, 2 and 3, respectively. Lanes 7, 8 and 9 contain Cataloochee samples 1, 2 and 3, respectively. Lanes 10, 11 and 12 contain Cataloochee samples 1, 2 and 3, respectively.

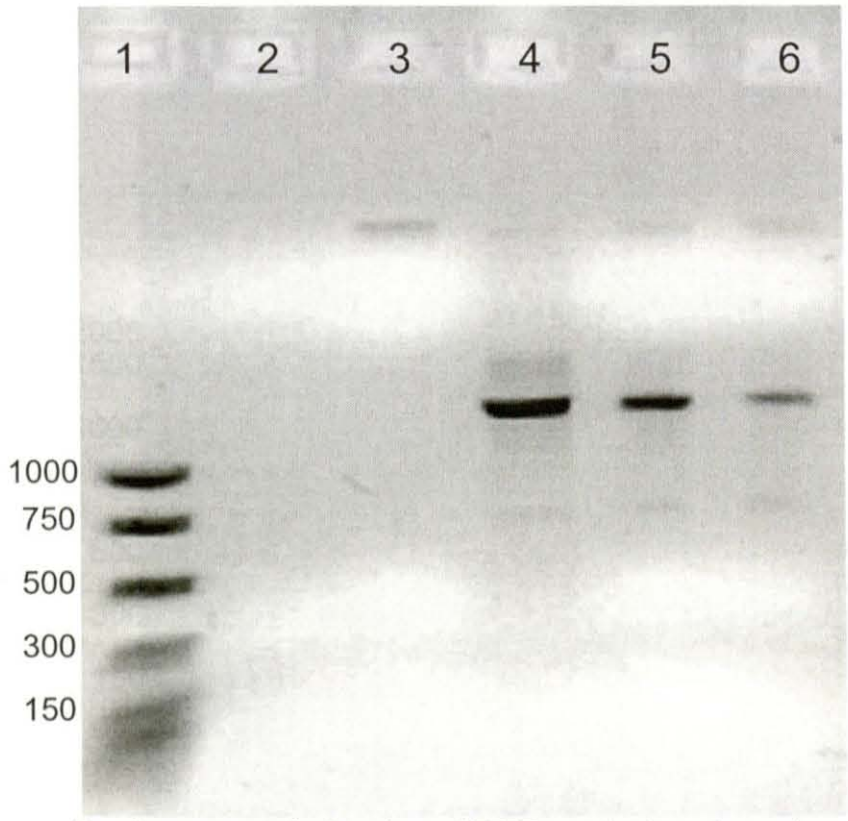

Figure 2a. Agarose gel of archaeal PCR products using primer set 46f and 1492r. Lane 1 contains the DNA ladder labeled in bp. Lanes 2 and 3 contain the negative and positive controls, respectively. Lane 4 contains Albright Grove sample 2, lane 5 contains Cataloochee sample 2 and lane 6 contains Purchase Knob sample 2.

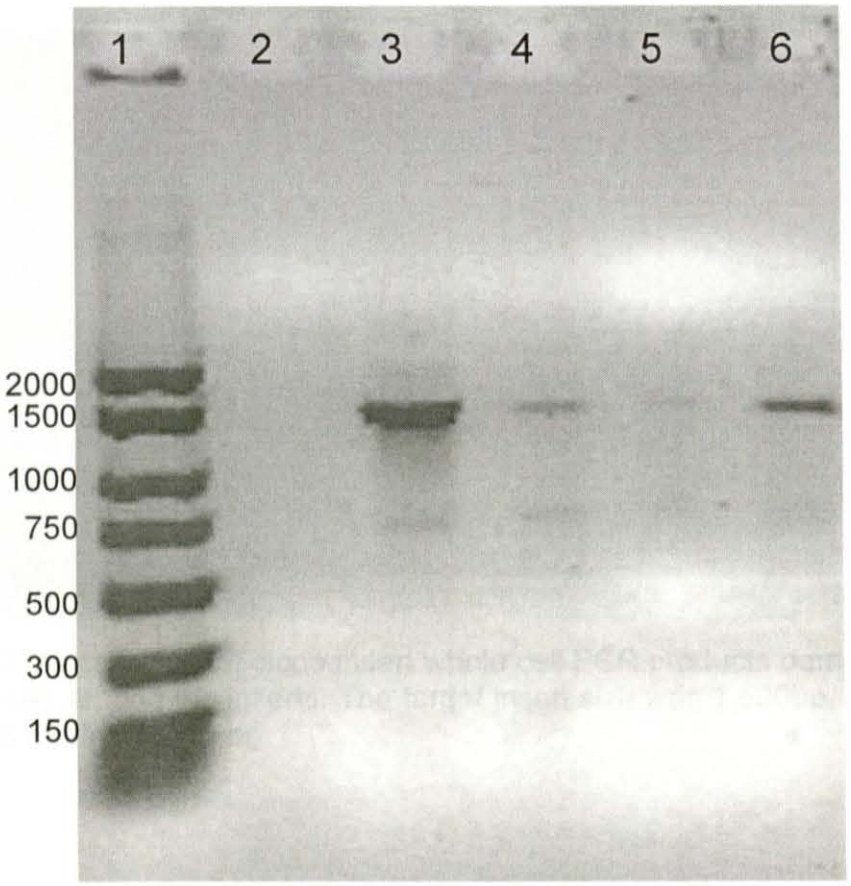

Figure 2b. Agarose gel of archaeal PCR products using primer set 21fa and 1492r. Lane 1 contains the DNA ladder labeled in bp. Lanes 2 and 3 contain the negative and positive controls, respectively. Lane 4 contains Albright Grove sample 1, lane 5 contains Cataloochee sample 2 and lane 6 contains Purchase Knob sample 3.

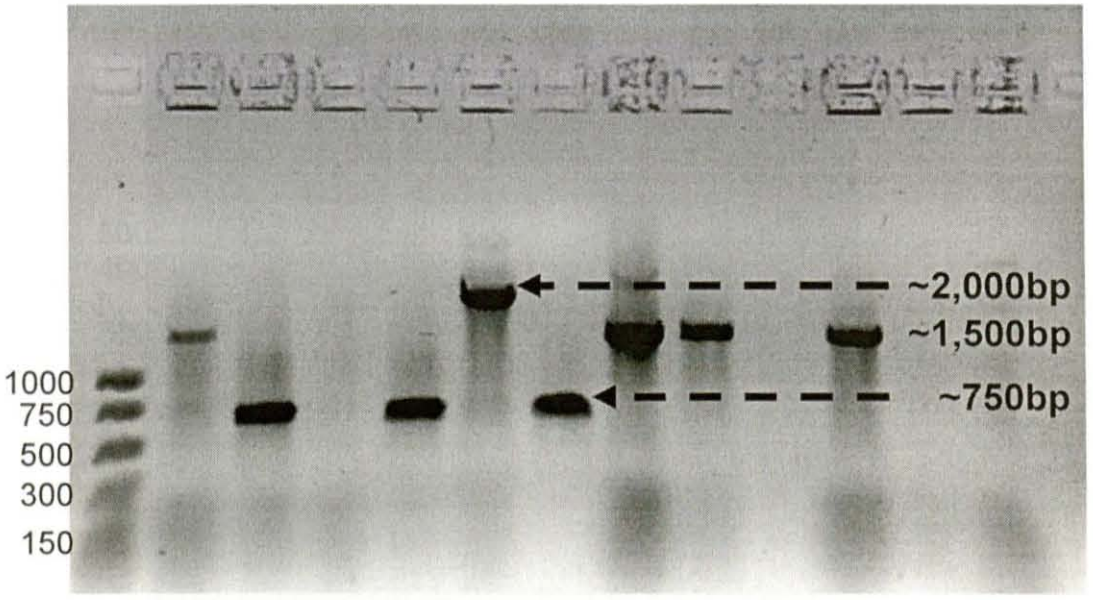

Figure 3. Agarose gel of clone insert whole cell PCR products demonstrating size differences among the inserts. The target insert size was 1,500bp. The DNA markers are labeled in bp

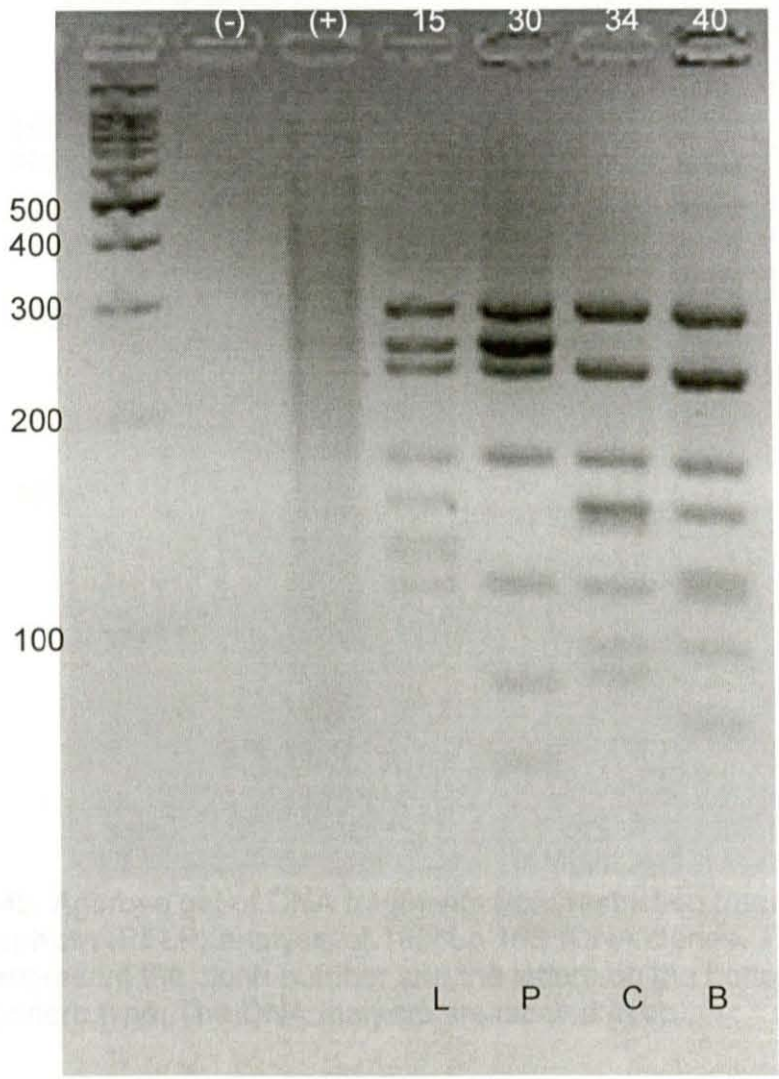

Figure 4a. Agarose gel of DNA fragments from restriction fragment length polymorphism (RFLP) analysis of 1500bp 16S rDNA clones. The numbers on top represent the clone number and the letters on the bottom represent the RFLP pattern type. The DNA markers are labeled in bp.

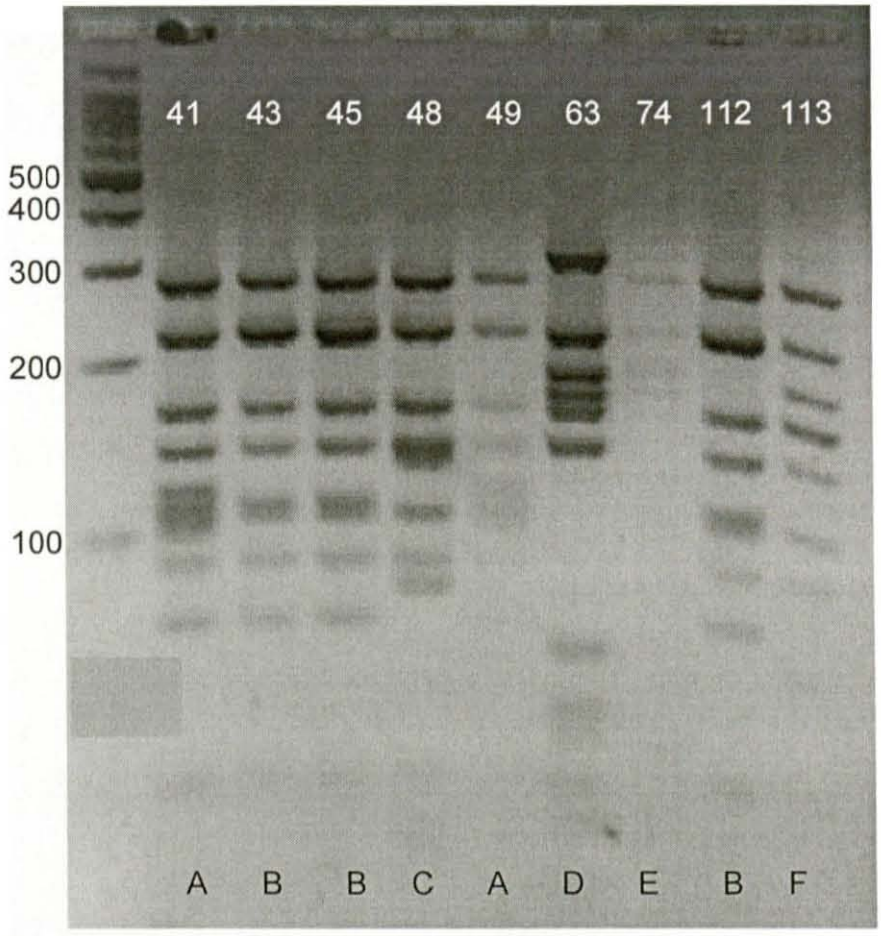

Figure 4b. Agarose gel of DNA fragments from restriction fragment length polymorphism (RFLP) analysis of 1500bp 16S rDNA clones. The numbers on top represent the clone number and the letters on the bottom represent the RFLP pattern type. The DNA markers are labeled in bp.

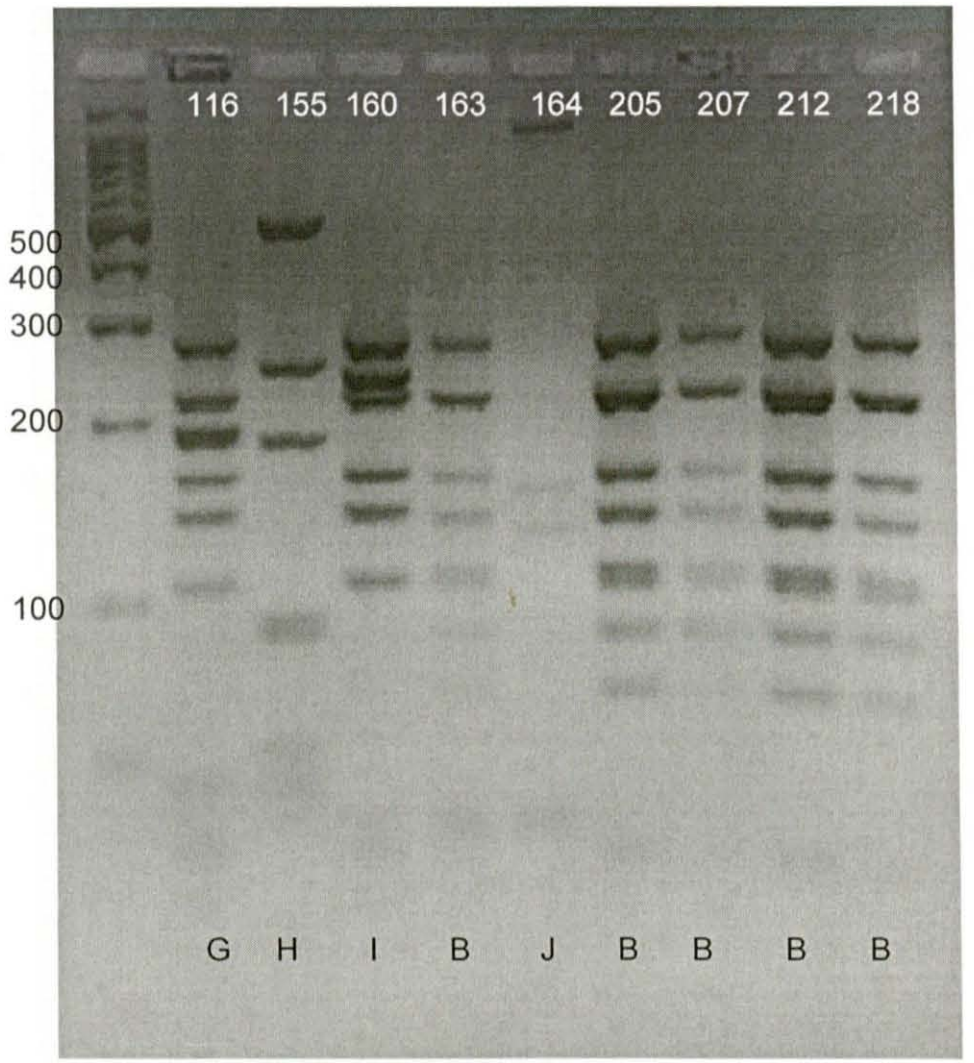

Figure 4c. Agarose gel of DNA fragments from restriction fragment length polymorphism (RFLP) analysis of 1500bp 16S rDNA clones. The numbers on top represent the clone number and the letters on the bottom represent the RFLP pattern type. DNA markers are labeled in bp.

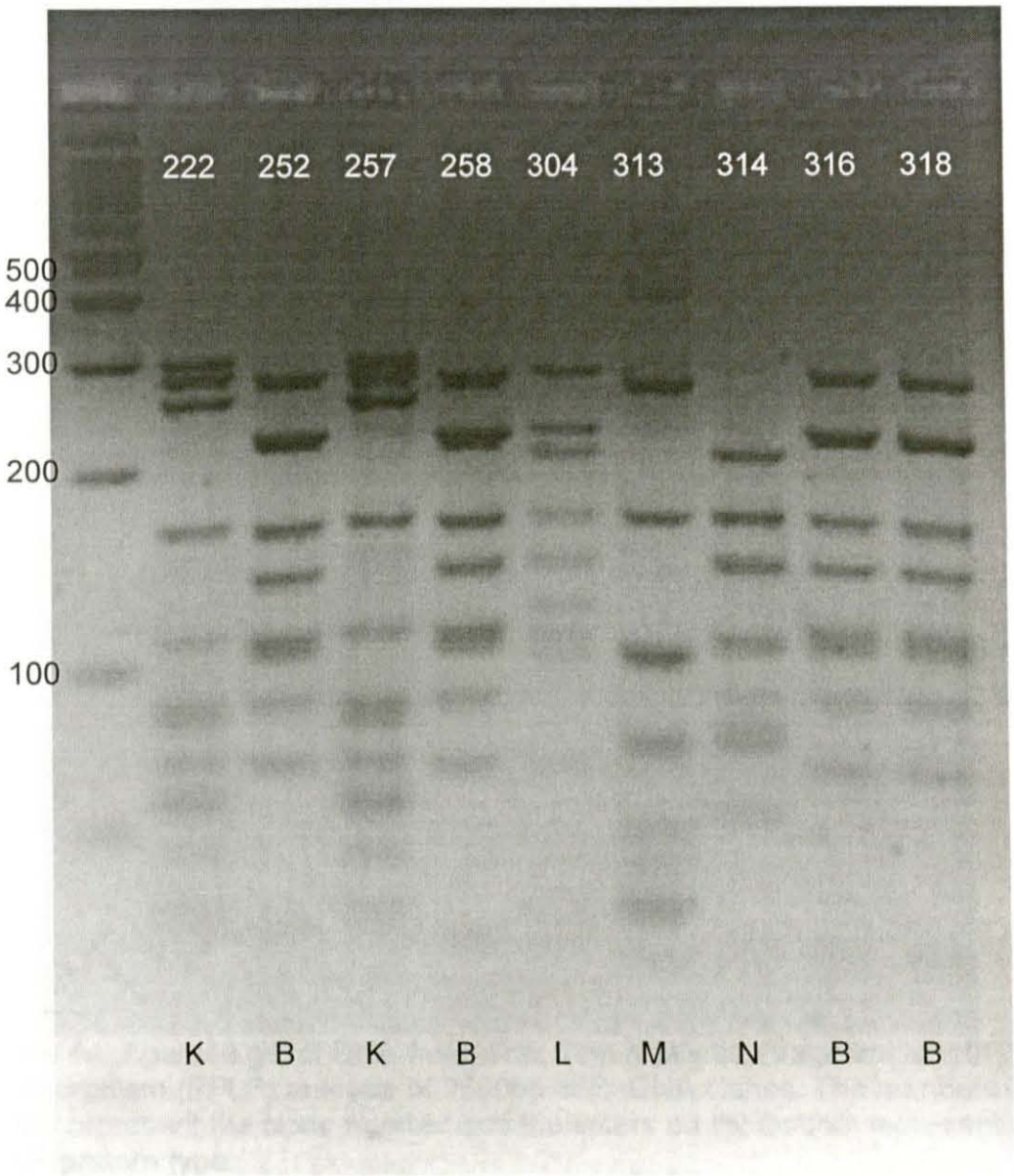

Figure 4d. Agarose gel of DNA fragments from restriction fragment length polymorphism (RFLP) analysis of 1500bp 16S rDNA clones. The numbers on top represent the clone number and the letters on the bottom represent the RFLP pattern type. DNA markers are labeled in bp.

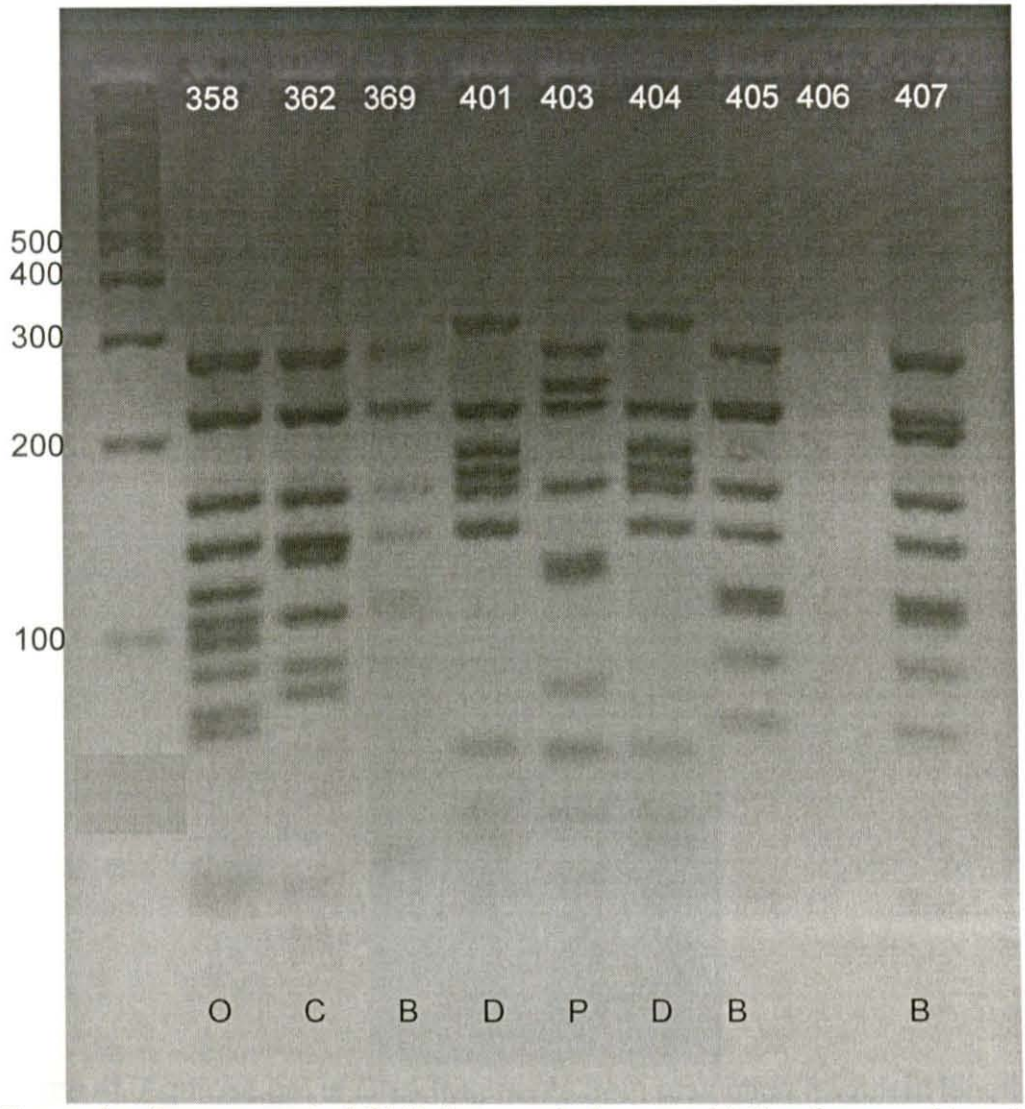

Figure 4e. Agarose gel of DNA fragments from restriction fragment length polymorphism (RFLP) analysis of 1500bp 16S rDNA clones. The numbers on top represent the clone number and the letters on the bottom represent the RFLP pattern type.

Figure 4f. Agarose gel of DNA fragments from restriction fragment length polymorphism (RFLP) analysis of 1500bp 16S rDNA clones. The numbers on top represent the clone number and the letters on the bottom represent the RFLP pattern type. Clone 61 was an insert that was smaller than the desired size after cloning and clone 62 was larger than the desired size after cloning. DNA ladder is labeled in bp.

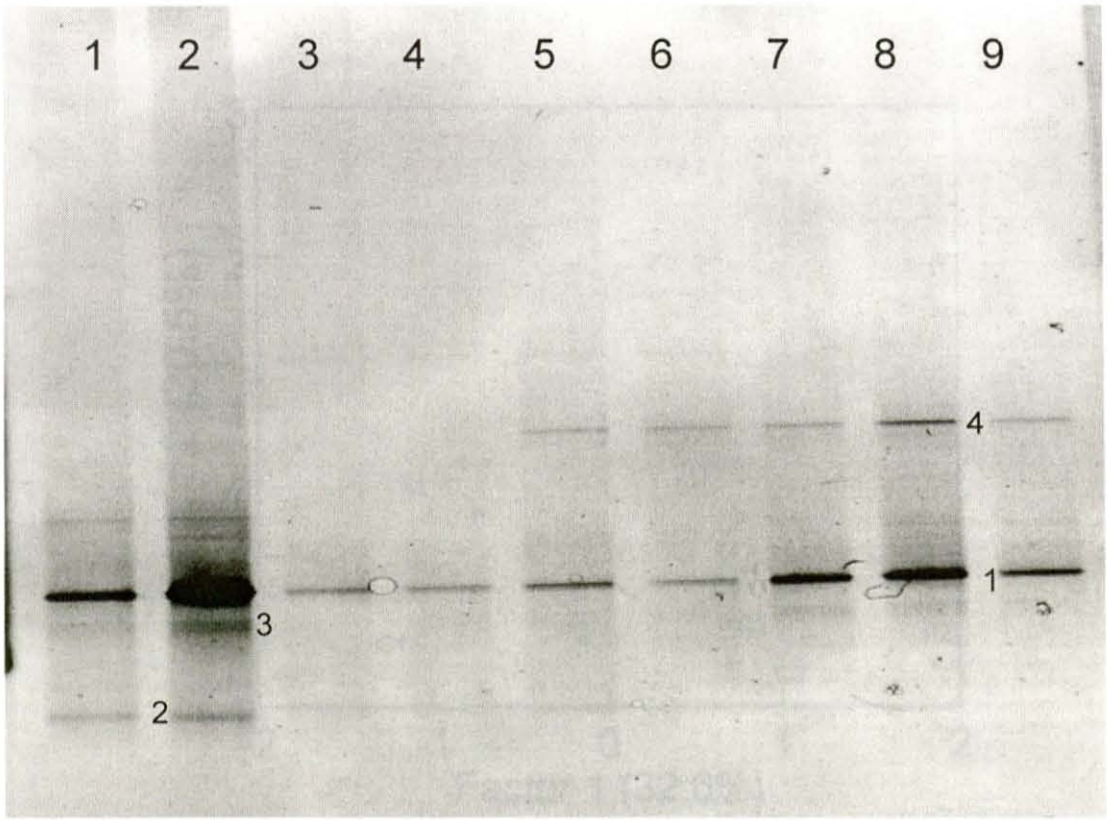

Figure 5. Denaturing gradient gel electrophoresis (DGGE) gel of archaeal community banding patterns from all soil sites. Excised bands are labeled to correspond with sequences DGGE 1-4. Lanes 1, 2 and 3 contain Albright Grove samples 1, 2 and 3, respectively, lanes 5, 6 and 7 contain Cataloochee samples 1, 2, and 3, respectfully and lanes 7, 8 and 9 contain Purchase Knob samples 1, 2 and 3 respectfully.

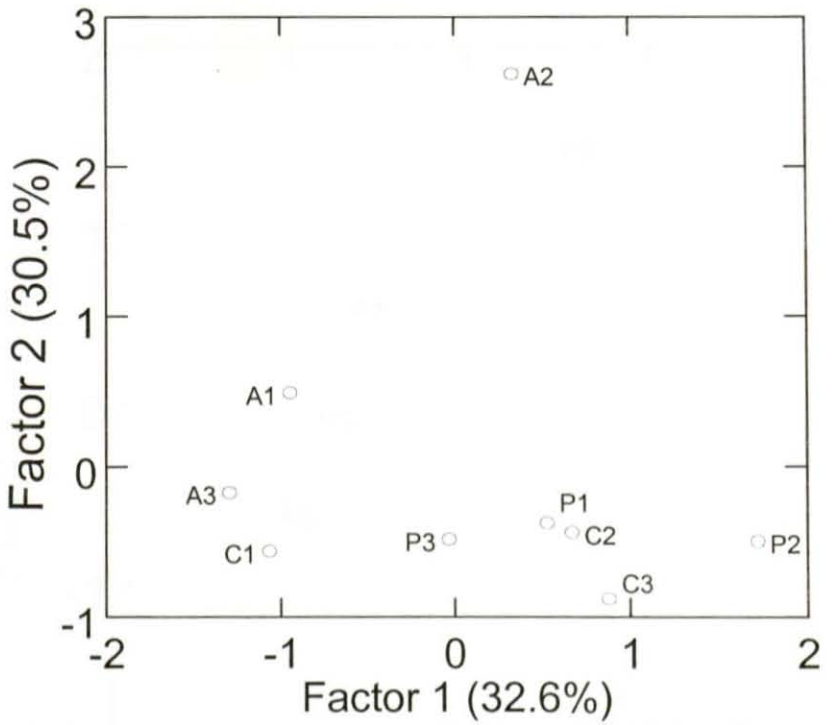

Figure 6a. Principal components analysis (PCA) of DGGE banding patterns for all nine soil samples. A, C and P represent Albright Grove, Cataloochee and Purchase Knob, respectively.

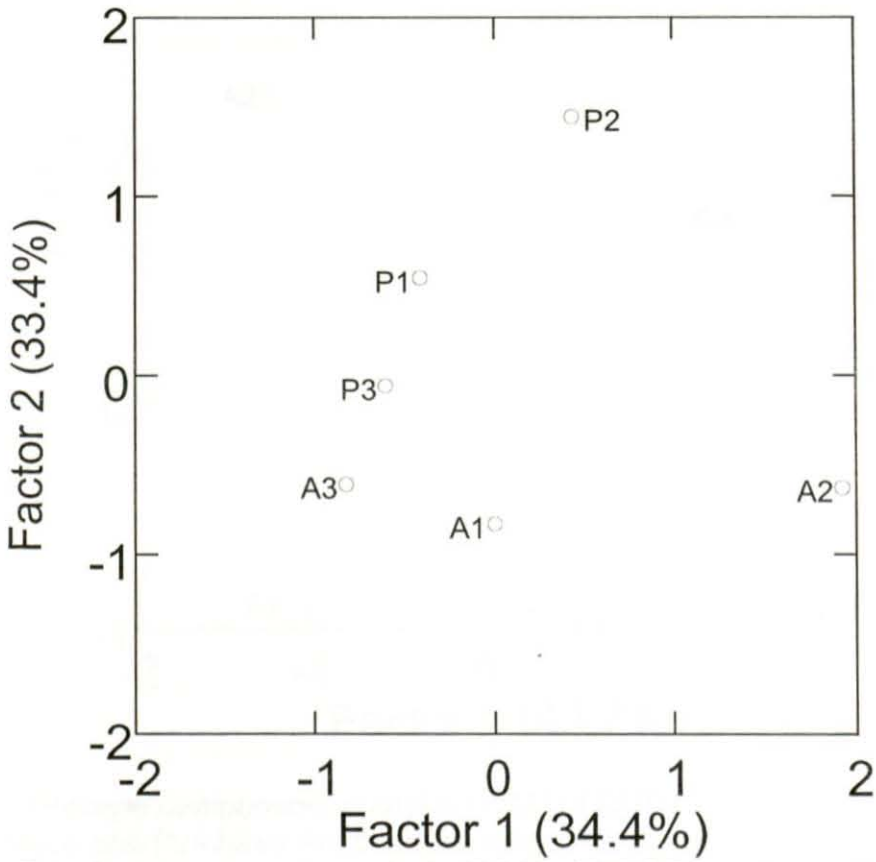

Figure 6b. Principal components analysis (PCA) of DGGE banding patterns for all Albright Grove and Cataloochee soil samples. A and P represent Albright Grove and Purchase Knob, respectively.

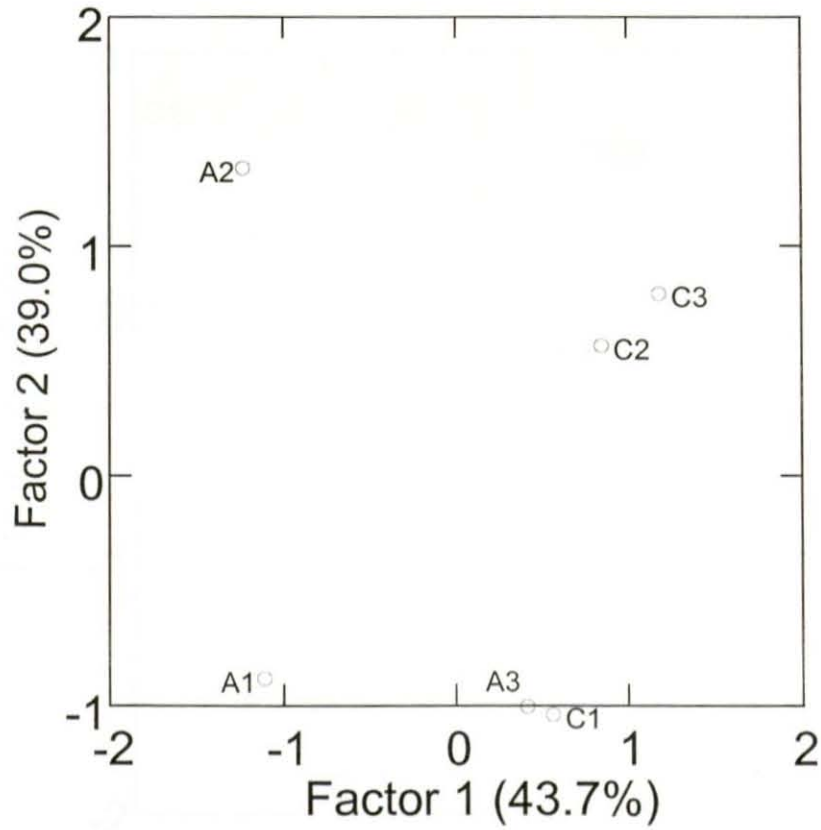

Figure 6c. Principle components analysis (PCA) of DGGE banding patterns for all Albright Grove and Purchase Knob soil samples. A and C represents Albright Grove and Cataloochee, respectively.

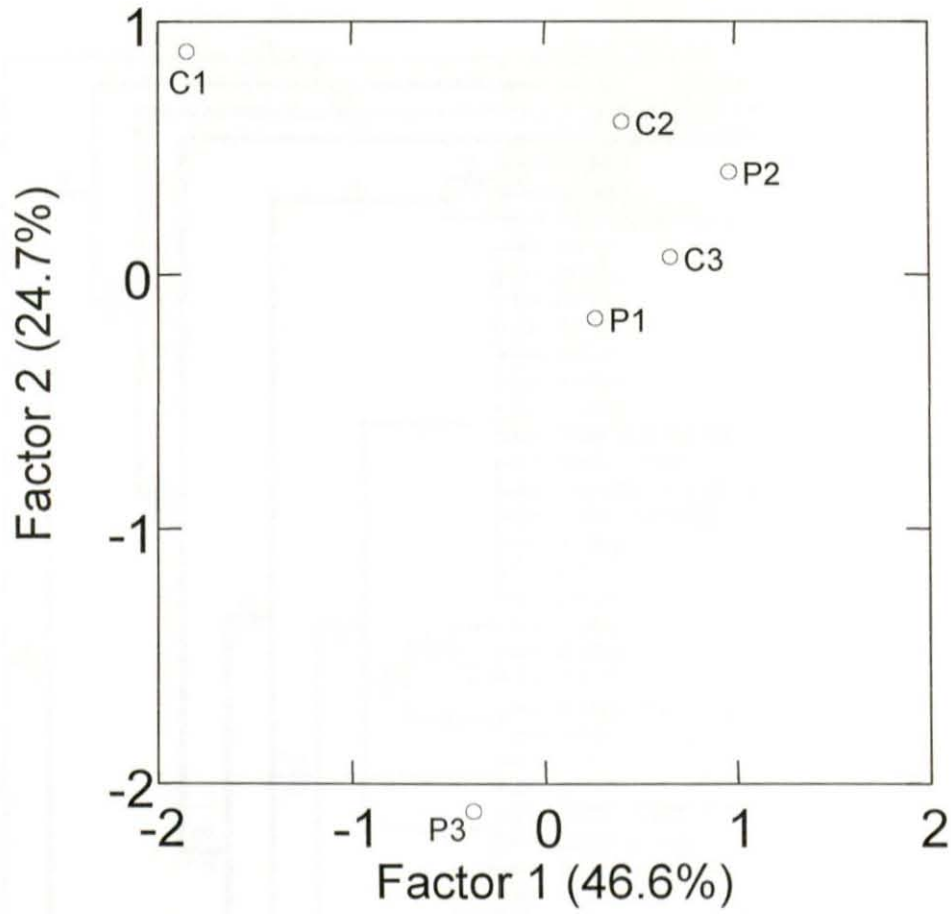

Figure 6d. Principal components analysis (PCA) of DGGE banding patterns for all Cataloochee and Purchase Knob soil samples. P and C represent Purchase Knob and Cataloochee, respectively.

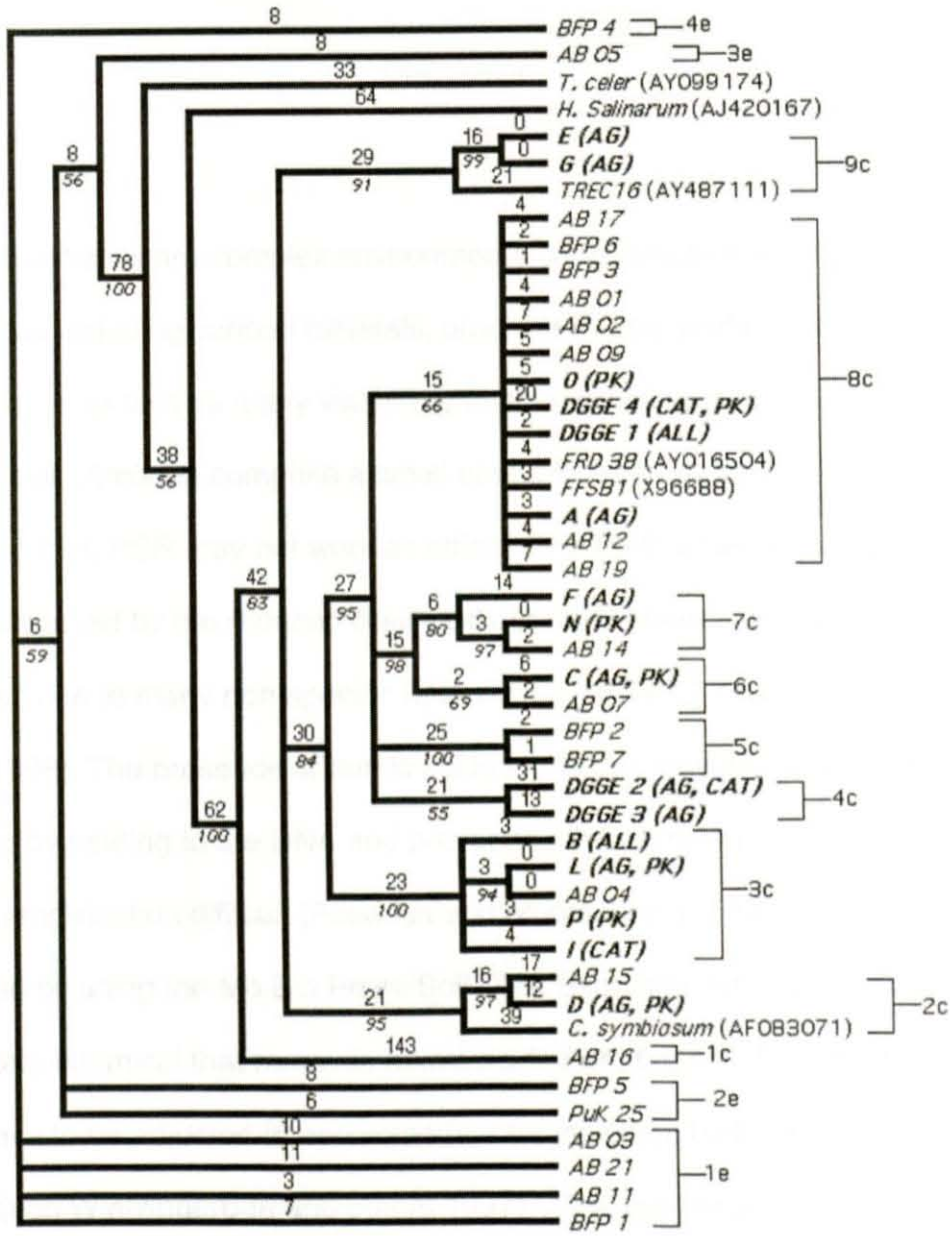

Figure 7. Unrooted phylogenetic tree of archaeal 16S rRNA gene sequences from soil samples from Great Smoky Mountains National Park. Bootstrap values are indicated by italicized type and branch lengths are indicated by normal type numbers. GenBank accession numbers are listed parenthetically. Unique clades are marked by brackets, a "c" next to the clade number represents sequences aligning to the phylum *Crenarchaea* and an "e" next to the clade number represents sequences aligning to the phylum *Euryarchaea*. Sequences obtained from this project are shown in boldface type. The clones sequenced from a GSMNP project performed in 2004 are indicated by the prefixes AB and PuK representing Alum Bluffs and Purchase Knob soil samples, respectively. BFP indicates soil archaeal clones from a project in 2003 at Beech Flats Prong.

Discussion

PCR

Forest soils are complex environments consisting of a variety of substances including various minerals, organics (humic acids, proteins, nucleic acids, etc...), as well as many viable organisms. When comparing the biomass that is in soil, *Archaea* comprise a small amount (Whitman and others 1998). As a result of that, PCR may not work as efficiently as with other organisms. This could be caused by the reduced likelihood of primers interacting with their templates due to many non-specific DNA molecules being present (Head and others 1998). The presence of humic acids in the soil may also lead to the inhibition by binding to the DNA and preventing the primers from binding, thus making amplification difficult (Friedrich and others 1997). This issue was alleviated by using the Mo Bio PowerSoil DNA Extraction kit, which uses a proprietary chemical that removes humic acids from the soil. PCR parameters often have to be adjusted to accommodate for inhibition by humics and other factors (von Wintzingerode and others 1997). All of the reagent concentrations were adjusted in this study to obtain high quality and quantity PCR products. No microbial diversity study involving soil will account for all species, so a balance between DNA extraction efficiency, PCR yield and sequence recovery is a tradeoff.

Molecular cloning

Cloning did not yield the targeted number of 450 clones among all three sites. This may have been caused by the low concentration of PCR products (Figure 2). This could be overcome by repeating the cloning, or concentrating the DNA before ligation. Compared to the number of sequences found when DGGE was performed directly after PCR, cloning did increase the number of species detected. In this study, there were 19 different cloned species versus 12 species from DGGE. A previous study also found 28 cloned species versus seven species when PCR products were directly used in DGGE (P. Drummond, unpublished data). It was also discovered that using 10µL of 50mg/mL X-gal did not work to accurately determine which colonies were truly white and which were blue. When the directed amount of 20µL of 50mg/mL X-gal was used it made the screening more accurate.

Variation in size of clone inserts

Whole cell PCR was used to amplify the insert from each clone. This technique revealed several different sizes of inserts (Figure 3). Since the fragments that were ligated into the plasmid came from the same PCR reaction, the sizes should have been nearly identical. If all of the inserts were the same size when they were ligated into the vectors, then any size difference was caused by the cloning. An attempt was made to sequence two of the unexpected size fragments, but they did not amplify prior to sequencing so all of the unexpected size fragments were not considered valid inserts.

Sequence analysis

Nearly all the sequences obtained in this study were related to *Archaea* that were discovered in other soil environments (Tables 3 and 4). GenBank and RDPII are dominated by bacterial data. RDPII is updated periodically with GenBank sequences but uses a different matching algorithm (Altschul and others 1997; Cole et al. 2003). Although both databases give a similarity value, they calculate them in different ways. GenBank uses the basic local alignment search tool (BLAST). BLAST finds the best match on a base by base comparison. The similarity value is given by dividing the number of bases that match between the query sequence and the subject sequence by the total number of bases queried. RDPII calculates the similarity value by identifying unique 7-oligomer stretches in the sequence and assigns each a number based on a numbering system. These numbers are compared to the database and the similarity values are calculated by dividing the number of unique oligomers shared between the query sequence and the subject sequence and dividing by the lowest number of unique oligomers in either of the two sequences.

The poor richness of archaeal sequences in the databases may give weak matches to what the query sequence may be closely related. The weak similarities may also be due to the sequences being fairly unique. There have not been any published studies that focused on *Archaea* in an environment that is similar to the soil of GSMNP. There have also not been any archaeal studies that were able to clearly identify metabolic processes from soil. This hampers our

ability to infer the possible metabolic or biochemical characteristics of the species and ultimately the community.

Although there is a poor richness of archaeal sequences in these databases, there could be some validity to the matches that were calculated by the algorithms. Many of the matches came from cold environments (boreal forest soil and tundra soil). The samples for this study and the Beech Flats Prong site were collected in the winter and at high elevations. The Beech Flats Prong samples were also stored at 4°C for three months before being analyzed. These conditions may have selected for an increase in the number of cold tolerant or requiring archaea, which would most likely be found in both boreal forest soil and tundra soil.

Phylogenetic analysis

The most obvious groupings in the phylogenetic tree is the distinction between the sequences that aligned with the phylum *Euryarchaea* and the phylum *Crenarchaea* (Figure 7). From the three sites that were investigated in this study, Albright Grove was the only site that had clones forming a unique clade (clade 9). Beech Flats Prong also had a clade (clade 5) that had clones that were unique to it. These clades could contain archaea that are adapted to those sites. DGGE analyses, although not producing as much species richness, did detect species that cloning did not. This may be due, in part, to using the 344F and 915R primers directly on the soil extracts versus using the 21Fa and 1492R primers. The differences in thermal cycler conditions may have also had

an effect. Although there are clades that are unique to a site, this is the minority. Clades 2, 3, 6, 7 and 8 each have clones from three or more locations in GSMNP. This may be an important finding because of the high overlap with so few clones. It seems archaea that live in soil may have broader distributions than soil bacteria. A similar study has been performed on the same samples, only the target was bacteria as opposed to archaea. The results showed that nearly all of the 180 RFLP patterns were unique with only a few overlaps between sites (M. Collins, unpublished data).

The "comb" effect seen in clade 8 was caused because when the tree was constructed, there were over 6,000 trees that had the same parsimony score and when a consensus tree was calculated, those sequences varied within that clade. As a result, the software placed them in the "comb," meaning that the exact placements of those sequences within that clade are not definite (K. Mathews, personal communication).

PCA Analysis

All of the PCA plots showed a separation of the Albright Grove site (Figures 5a, b, c), but did not separate Cataloochee from Purchase Knob (Figure 5d). This can be partially explained by the number of DGGE bands. Samples C2, C3, P1 and P3 consistently were grouped together. This could be explained since C2, C3, P1 and P2 had very similar numbers of DGGE bands and they shared a unique species. The location of P3 seems to be an anomaly since it has a much lower number of DGGE bands (2 bands) from the rest of the samples

and does not share in any unique bands. Samples A3 and C1 both share the same banding pattern, including a unique band and are located close to each other. Sample A2 had nine DGGE bands and two unique bands. Sample P2 had seven DGGE bands and one unique band. This may explain why A2 and P2 were consistently separated from the rest.

Ecological role of archaea

Due to the lack of cultured archaea, especially non-thermophilic *Crenarchaea*, knowing the roles the organisms from the study play *in situ* is virtually impossible. Ribosomal sequences alone are not good predictors of ecological function (Madsen 1998). The only guidance that is available is from information that has been collected from the very few cultured non-thermophilic *Crenarchaea* as well as other *Archaea* (both *Crenarchaea* and *Euryarchaea*).

The known metabolisms found within *Archaea* are largely from extreme organisms. Halophilic *Archaea* are heterotrophic and use amino acids or organic acids as their main energy source (Madigan and others 2003). Thermophilic *Archaea* have a variety of metabolic processes. Many thermophilic *Crenarchaea* depend on sulfur as their electron donor. Many can reduce S^0, SO_4^{2-}, $S_2O_3^{2-}$ or SO_3^{2-} and oxidize organic compounds, producing H_2S. Others can oxidize S^0 and Fe^{2+} aerobically producing H_2SO_4 or $FeSO_4$; or they can anaerobically reduce NO_3^- or Fe^{3+} yielding NO_2^- or Fe^{2+}. Certain marine archaea have the ability to use a protein known as bacteriorhodopsin (Béjà and others 2000). Bacteriorhodopsin is a trans-membrane protein that can absorb light and catalyze a proton motive

force which drives ATPase producing energy for the cell in the form of ATP. Another common metabolism for archaea is methanogenesis. Methanogenesis is performed in many ways by archaea from the phylum *Euryarchaea*. A few pathways include using inorganic carbon (CO_2) as their only carbon source. They combine CO_2 with H_2 to produce methane. Others can use $HCOO^-$ and organic molecules with H_2. Others only need CH_3OH to produce methane (Madigan and others 2003). These metabolisms are strictly anaerobic. Archaea that are related closest to methanogens have been found in GSMNP soil in previous studies (G. Parise, unpublished data; P. Drummond, unpublished data).

There has yet to be a non-thermophilic crenarchaeon from a soil environment obtained in pure culture, so there are no definitive metabolic processes from these organisms. The closest study came from Simon, et al. (2005) where non-thermophilic *Crenarchaea* were grown in enrichment cultures from tomato plant roots, meaning they were not the only organisms in the culture. They found that the *Crenarchaea* only grew in media that contained the plant root extract. There has been more work done on non-thermophilic *Crenarchaea* in marine water and gaining insights on the metabolic properties of those *Crenarchaea* were more successful. The only non-thermophilic *Crenarchaea* grown thus far in pure culture is a crenarchaeon from ocean sediments (Könneke and others 2005). The authors discovered that this organism oxidizes ammonia into nitrite and fixes CO_2. Other studies have described *Cenarchaeum symbiosum*, a crenarchaeon that is a symbiont to the marine sponge, *Axinella*

mexicana (Preston and others 1996). It too was found to oxidize ammonia. It has been shown through metagenomic studies that some soil *Crenarchaea* likely have the ability to oxidize ammonia (Nicol and Schleper 2006). Metagenomic studies involve cloning large stretches of DNA using the 16S rDNA as an "anchor" (Handelsman 2004). These contigs can be arranged using overlapping sequences and then the resulting sequence can be searched against sequence databases for homologues of known proteins. These methods allowed the non-thermophilic marine ammonia oxidizing crenarchaeon to be cultured since it was known that the organism contained the enzymes used to oxidize ammonia.

It is currently not known what the *Archaea* are doing in the soils of GSMNP, but it is possible to infer by using the data from other studies that at least some of them are oxidizing ammonia. It seems that sequencing more genes from these organisms and comparing them to known genes using metagenomics can allow for insight to the metabolic activities of these species and ultimately confirm the biochemical processes by culturing them.

Conclusions and Future Work

Although the desired number of archaeal clones (450) was not attained in this study, molecular cloning was observed to increase the number of detectable sequences in a sample when contrasted with DGGE. There were 19 unique cloned sequences as compared to 12 DGGE bands. The DGGE results suggest that perhaps environmental disturbance may have an effect many years after the disturbance, since there was a separation between Albright Grove, the undisturbed site, and the two disturbed sites and there was not a clear separation between the two disturbed sites.

GSMNP continues to provide archaeal 16S rDNA sequences that appear to be quite unique when compared to the currently available sequences. These organisms quite possibly may perform novel biochemical processes that could have an impact on the way forests function.

Future studies could include using the sequences generated by this study to design specific oligonucleotide primers with fluorescent tags bound to them to probe the original samples in conjunction with hybridization and fluorescent microscopy to examine their abundance and morphological characteristics in soil (Amann and others 1992).

References

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z, Miller, W. and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research. 25:3389-402.

Amann, R.I., Zarda, B., Stahl, D.A. and Schleifer, K-H.. 1992. Identification of individual prokaryotic cells by using enzyme-labeled, rRNA-targeted oligonucleotide probes. Applied and Environmental Microbiology 58:3007-11.

Béjà, O., Arvind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P., Jovanovich, S.B., Gates, C.M., Feldman, R.A., Spudich, J.L., Spudich, E.N. and DeLong, E.F. 2000. Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. Science. 289:1902-6.

Bintrim, S.B., Donohue, T.J., Handelsman, J., Roberts, G.P. and Goodman, R.M. 1997. Molecular phylogeny of archaea from soil. Proceedings of the National Academy of Sciences of the United States of America. 94:277-82.

Bond, P.L., Smriga, S.P. and Banfield, J.F. 2000. Phylogeny of microorganisms populating a thick, subaerial, predominantly lithotrophic biofilm at an extreme acid mine drainage site. Applied and Environmental Microbiology. 66:3842-9.

Casamayor, E.O., Massana, R., Benlloch, S., Øverås, L., Diez, B., Goddard, V.J., Gasol, J.M., Joint, I., Rodríguez-Valera, F. and Pedrós-Alió, C. 2002. Changes in archaeal, bacterial and eukaryal assemblages along a salinity gradient by comparison of genetic fingerprinting methods in a multipond solar saltern. Environmental Microbiology. 4:338-48.

Cole, J.R., Chai, B., Marsh, T.L., Farris, R.J., Wang, Q., Kulam, S.A., Chandra, S., McGarrell, D.M., Schmidt, T.M., Garrity, G.M. and Tiedje, J.M. 2003. The Ribosomal Database Project (RDP-II): Previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. Nucleic Acids Research. 31:442-3.

Copeland, H.F. 1938. The kingdoms of organisms. Quarterly Review of Biology. 13:383-420.

Delong, E.F. 1992. Archaea in coastal marine environments. Proceedings of the National Academy of Sciences of the United States of America. 89:5685-9.

DeLong E.F., Wickham, G.S. and Pace, N.R. 1989. Phylogenetic stains: Ribosomal RNA-based probes for the identification of single cells. Science. 243:1360-3.

Discover Life in America. Great Smoky Mountains National Park All Taxa Biodiversity Inventory. Available from: http://www.dlia.org. Accessed October 2005.

Fredrickson, J.K., Zachara, J.M., Balkwill, D.L., Kennedy, D., Li, S.W., Kostandarithes, H.M., Daly, M.J., Romine M.F. and Brockman, F.J. 2004. Geomicrobiology of high-level nuclear waste-contaminated vadose sediments at the Hanford site, Washington State. Applied and Environmental Microbiology. 70:4230-41.

Handelsman, J. 2004. Metagenomics: Application of genomics to uncultured microorganisms. Microbiology and Molecular Biology Reviews. 68:669-85.

Head, I.M., Saunders, J.R., and Pickup, R.W. 1998. Microbial evolution, diversity, and ecology: a decade of ribosomal RNA analysis of uncultivated microorganisms. Microbial Ecology. 35:1-21.

Huber, H., Hohn, M.J., Rachel, R., Fuchs, T., Wimmer, V.C. and Stetter, K.O. 2002. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. Nature. 417:63–7.

Jurgens, G., Lindstrom, K. and Saano, A. 1997. Novel group within the kingdom *Crenarchaeota* from boreal forest soil. Applied and Environmental Microbiology. 63:803-5.

Jurgens, G.N. and Saano, A. 1999. Diversity of soil archaea in boreal forest before, and after clear-cutting and prescribed burning. FEMS Microbiology Ecology. 29:205-13.

Kaplan, C.W., Astaire, J.C., Sanders, M.E., Reddy, B.S. and Kitts, C.L. 2001. 16S ribosomal DNA terminal restriction fragment pattern analysis of bacterial communities in feces of rats fed *Lactobacillus acidophilus* NCFM. Applied and Environmental Microbiology. 67:1935-9.

Knittel, K., Lösekann, T., Boetius, A., Kort, R. and Amann, R. 2005. Diversity and distribution of methanotrophic archaea at cold seeps. Applied and Environmental Microbiology. 71:467-79.

Kobs, G. 1997. Cloning blunt-end DNA fragments into the pGEM®-T Vector Systems. Promega Notes 62:15-8.

Könneke, M., Bernhard, A.E., De La Torre, J.R., Walker, C.B., Waterbury, J.B. and Stahl, D.A. 2005. Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature. 437:543-6

Konopka, A., Zakharova, T., Bischoff, M., Oliver, L., Nakatsu, C. and Turco, R.F. 1999. Microbial biomass and activity in lead-contaminated soil. Applied and Environmental Microbiology. 1999. 65:2256-9.

Lane, D.J. 1991. 16S/23S rRNA sequencing, p. 115-175. *In* Stackebrandt, E. and Goodfellow, M. (ed.), Nucleic Acid Techniques in Bacterial Systematics. John Wiley & Sons, Chichester, England.

Lepage, E., Marguet, E., Geslin, C., Matte-Tailliez, O., Zillig, W., Forterre, P. and Tailliez, P. 2004. Molecular diversity of new *Thermococcales* isolates from a single area of hydrothermal deep-sea vents as revealed by randomly amplified polymorphic DNA fingerprinting and 16S rRNA gene sequence analysis. Applied and Environmental Microbiology. 70:1277-86.

Madigan, M.T., Martinko, J.M. and Parker, J. 2003. Brock Biology of Microorganisms. Pearson Educations, Upper Saddle River, NJ, ed. 10.

Madsen, E.L. 1998. Epistemology of environmental microbiology. Environmental Science and Technology. 32:429-39.

Miller, D.N., Bryant, J.E., Madsen, E.L. and Ghiorse, W.C. 1999. Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. Applied and Environmental Microbiology. 65:4715-24.

Muyzer, G. and Smalla, K. 1998. Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. Antonie Van Leeuwenhoek. 73:127-41.

National Park Service. Great Smoky Mountains National Park. Available from: http://www.nps.gov/grsm/gsmsite/welcome.html. Accessed. October 2005.

Nicol, G.W. and Schleper, C. 2006. Ammonia-oxidising Crenarchaeota: important players in the nitrogen cycle? Trends in Microbiology. 14:207-12.

Nicol, G.W., Tscherko, D., Embley, T.M. and Prosser, J.I. 2005. Primary succession of soil Crenarchaeota across a receding glacier Foreland. Environmental Microbiology. 7:337-47.

O'Connell, S.P., Lehman, R.M., Snoeyenbos-West, O., Winston V.D., Cummings D.E., Watwood, M.E. and Colwell, F.S. 2003. Detection of *Euryarchaeota* and *Crenarchaeota* in an oxic basalt aquifer. FEMS Microbiology Ecology. 44:165-73.

Preston, C.M., Wu, K.Y., Molinski, T.F. and DeLong, E.F. 1996. A psychrophilic crenarchaeon inhabits a marine sponge: *Cenarchaeum symbiosum* gen. nov. Proceedings of the National Academy of Sciences of the United States of America. 93:6241-6.

Qiu, X., Wu, L., Huang, H., McDonel, P.E., Palumbo, A.V., Tiedje, J.M. and Zhou, J. 2001. Evaluation of PCR-generated chimeras, mutations and heteroduplexes with 16S rRNA gene-based cloning. Applied and Environmental Microbiology. 67:880-7.

Reeve, J.N. 1999. *Archaebacteria* Then ... Archaes now (are there really no archaeal pathogens?). Journal of Bacteriology. 181:3613–7.

Sharkey, M.J. 2001. The All Taxa Biological Inventory of the Great Smoky Mountains National Park. Florida Entomologist. 84:556-64.

Stahl, D.A., Lane, D.J., Olsen, G.J. and Pace, N.R. 1985. Characterization of a Yellowstone hot spring microbial community by 5S rRNA sequences. Applied and Environmental Microbiology. 49:1379-84.

Stainer, R.Y. and Van Niel, C.B. 1941. The main outlines of bacterial classification. Journal of Bacteriology. 42:437-66.

Takai, K., Moser, D.P., DeFlaun, M., Onstott, T.C. and Fredrickson, J.K. 2001. Archaeal diversity in waters from deep South African gold mines. Applied and Environmental Microbiology. 67:5750-60.

Teira, E., T. Reinthaler, A. Pernthaler, J. Pernthaler, and G. J. Herndl. 2004. Combining catalyzed reporter deposition-fluorescence in situ hybridization and microautoradiography to detect substrate utilization by bacteria and archaea in the deep ocean. Applied and Environmental Microbiology. 70:4411-4.

Torsvik, V., Goksoyr, J. and Daae, F.L. 1990. High diversity in DNA of soil bacteria. Applied and Environmental Microbiology. 56:782-7.

Vetriani, C., Jannasch, H.W., MacGregor, B.J., Stahl, D.A. and Reysenbach, A.L. 1999. Population structure and phylogenetic characterization of marine benthic archaea in deep-dea sediments. Applied and Environmental Microbiology. 65:4375-84.

Whitman, W.B., Coleman, D.C. and Wiebe, W.J. 1998. Prokaryotes: the unseen majority. Proceedings of the National Academy of Sciences of the United States of America. 95:6578-83.

Whittaker, R.H. 1969. New concepts of kingdoms of organisms. Science. 163:150-60.

Willis, D.K. and Goodman, R.M. 2005. Cultivation of mesophilic soil crenarchaeotes in enrichment cultures from plant roots. Applied and Environmental Microbiology. 71:4751-60.

von Wintzingerode, F., Göbel, U.B. and Stackebrandt, E. 1997. Determination of microbial diversity in environmental samples: Pitfalls of PCR-based rRNA analysis. FEMS Microbiology Reviews 21:213-29.

Woese, C.R. 1987. Bacterial evolution. Microbiological Reviews. 51:221-71.

Woese, C.R. 1994. There must be a prokaryote somewhere: Microbiology's search for itself. Microbiological Reviews. 58:1-9.

Woese, C.R. and Fox, G.E. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proceedings of the National Academy of Sciences of the United States of America. 74:5088-90.

Woese, C.R., Kandler, O. and Wheelis, M.L. 1990. Towards a natural system of organisms: Proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. Proceedings of the National Academy of Sciences of the United States of America. 87:4576-9.

Appendix

Appendix 1

Sequence similarity table comparing all of the sequences used in Figure 2. Values indicate percent similarity. Values above 90% are shown in boldface.

	L	B	A	C	D	E	F	G	I	N	O	P	BFP 1	BFP 2
L	100	**99**	88	88	84	84	87	84	**99**	88	88	**99**	72	87
B		100	89	88	84	84	88	84	**99**	88	88	**99**	72	87
A			100	**93**	82	85	92	85	88	93	**98**	89	69	**92**
C				100	83	84	95	84	88	**97**	**93**	88	71	**91**
D					100	86	81	86	84	83	83	84	72	82
E						100	84	**100**	84	84	84	84	72	84
F							100	84	88	**97**	**92**	88	70	**91**
G								100	84	84	84	84	72	84
I									100	88	88	**99**	72	87
N										100	**93**	88	69	**91**
O											100	88	69	**92**
P												100	72	87
BFP 1													100	71
BFP 2														100
BFP 3														
BFP 4														
BFP 5														
BFP 6														
BFP 7														
DGGE 1														
DGGE 2														
DGGE 3														
DGGE 4														
AB 01														
AB 02														
AB 03														
AB 04														
AB 05														
AB 07														
AB 09														
AB 11														
AB 12														
AB 14														
AB 15														
AB 16														
AB 17														
AB 19														
AB 21														
PuK 25														
C. symbiosum														
FFSB1														
FRD38														
H. salinarum														
T. celer														
TREC16														

(cont'd). Sequence similarity table comparing all of the sequences used in Figure 2. Values indicate percent similarity. Values above 90% are shown in boldface.

	BFP 3	BFP 4	BFP 5	BFP 6	BFP 7	DGGE 1	DGGE 2	DGGE 3	DGGE 4	AB 01	AB 02	AB 03	AB 04	AB 05	AB 07	AB 09
L	88	72	72	88	87	89	85	86	82	88	89	72	**99.8**	71	88	88
B	88	71	72	88	87	89	85	86	82	88	89	71	**99**	70	88	88
A	**99**	68	68	**99**	**92**	**99.8**	87	**92**	88	**99**	**99**	67	88	66	**93**	**98**
C	**94**	69	69	**93**	**91**	**93**	**90**	88	83	**93**	**93**	68	88	68	**98**	**93**
D	83	71	72	83	83	83	78	78	76	83	82	70	84	70	82	82
E	85	71	72	85	85	85	79	81	79	84	85	70	84	70	84	83
F	**92**	68	69	**92**	**91**	**92**	88	87	82	**92**	**92**	69	87	69	**96**	**92**
G	85	71	72	85	85	85	79	81	79	84	85	70	84	70	84	83
I	88	72	72	88	87	88	84	86	82	88	88	72	**98**	71	88	87
N	**94**	68	68	**93**	**92**	**93**	89	89	83	**93**	**93**	68	87	68	**97**	**93**
O	**99**	68	68	**99**	**93**	**99**	87	**92**	88	**99**	**98**	67	87	66	**93**	**98**
P	88	71	72	88	87	89	84	86	82	88	89	71	**98**	70	88	88
BFP 1	69	**97**	96	69	71	69	67	66	64	69	69	**96**	72	95	70	69
BFP 2	**92**	69	69	**92**	**99**	**92**	85	87	82	**92**	92	69	87	68	**91**	**91**
BFP 3	100	68	68	**99**	**93**	**99**	87	**91**	88	**99**	**98**	67	88	66	**94**	**98**
BFP 4		100	**96**	69	69	68	65	66	64	69	68	**97**	72	**95**	68	68
BFP 5			100	68	69	68	65	66	64	69	68	**95**	72	**95**	68	68
BFP 6				100	**93**	**99**	87	**92**	88	**99**	**99**	67	88	67	**94**	**98**
BFP 7					100	**93**	85	88	83	**93**	**92**	69	87	68	**92**	**92**
DGGE 1						100	87	**92**	88	**99**	**99**	67	88	66	**94**	**98**
DGGE 2							100	**90**	82	87	87	64	84	64	**90**	87
DGGE 3								100	87	**92**	92	65	86	64	88	**91**
DGGE 4									100	88	88	62	82	62	83	87
AB 01										100	**99**	67	88	67	**94**	**99**
AB 02											100	67	88	67	**94**	**99**
AB 03												100	71	**95**	68	67
AB 04													100	71	88	87
AB 05														100	67	66
AB 07															100	**94**
AB 09																100
AB 11																
AB 12																
AB 14																
AB 15																
AB 16																
AB 17																
AB 19																
AB 21																
PuK 25																
C. symbiosum																
FFSB1																
FRD38																
H. salinarum																
T. celer																
TREC16																

(cont'd). Sequence similarity table comparing all of the sequences used in Figure 2. Values indicate percent similarity. Values above 90% are shown in boldface.

	AB 11	AB 12	AB 14	AB 15	AB 16	AB 17	AB 19	AB 21	PuK 25	*C. symbiosum*	FFSB1	FRD38	*H. salinarum*	*T. celer*	TREC16
L	73	88	87	85	59	88	88	71	72	80	88	88	68	76	82
B	72	88	88	84	58	89	88	70	72	80	88	88	67	76	82
A	68	**99**	**92**	81	58	**99**	**98**	68	68	78	**99**	**99**	67	73	83
C	70	**93**	**96**	82	58	**94**	**93**	68	70	79	**93**	**93**	68	72	83
D	71	82	83	**94**	59	83	82	70	72	87	82	82	67	72	87
E	70	85	84	85	65	84	85	70	71	81	84	85	69	74	**93**
F	69	**91**	**96**	80	57	**92**	**93**	68	69	78	**92**	**91**	68	73	82
G	70	85	84	85	65	84	85	70	71	81	84	85	69	74	**93**
I	73	88	87	83	59	88	88	71	72	80	88	88	67	76	81
N	68	**92**	**99.6**	82	57	**93**	**94**	67	68	80	**93**	**92**	67	73	83
O	68	**98**	**93**	81	58	**98**	**98**	68	68	78	**98**	**98**	67	73	82
P	72	88	88	84	59	89	88	70	72	80	88	88	67	76	82
BFP 1	**98**	69	69	72	53	69	68	**96**	**97**	69	69	69	70	78	69
BFP 2	69	**91**	**91**	82	58	**92**	**92**	69	69	79	**92**	**91**	66	73	81
BFP 3	68	**99**	**93**	82	58	**99**	**98**	68	68	78	**99**	**99**	68	73	83
BFP 4	**98**	68	67	72	52	68	67	**97**	**97**	69	68	68	70	77	68
BFP 5	97	68	67	72	53	68	68	**95**	**97**	69	68	68	70	77	69
BFP 6	68	**99**	**93**	82	58	**99**	**98**	68	68	78	**99**	**99**	67	73	82
BFP 7	69	**92**	**91**	83	58	**92**	**93**	69	69	80	**92**	**92**	67	73	81
DGGE 1	68	**99.6**	**92**	81	59	**99**	**98**	68	68	79	**99**	**99.6**	68	74	83
DGGE 2	65	86	88	77	53	88	87	64	66	75	87	86	64	70	77
DGGE 3	66	**92**	88	78	54	**91**	**92**	65	66	76	**91**	**92**	65	71	79
DGGE 4	64	88	83	75	53	88	87	63	64	72	88	88	62	68	76
AB 01	68	**98**	**93**	82	58	**98**	**99**	68	68	78	**99**	**98**	68	74	82
AB 02	68	**99**	**93**	81	58	**98**	**98**	68	68	78	**98**	**99**	67	74	83
AB 03	**98**	67	68	71	52	67	66	**97**	**96**	68	67	67	69	77	68
AB 04	72	88	87	84	59	88	87	71	72	80	87	88	67	76	81
AB 05	**96**	66	67	72	52	66	66	**96**	**96**	67	66	66	70	77	68
AB 07	69	**93**	**97**	82	58	**94**	**94**	68	69	80	**93**	**93**	68	73	82
AB 09	68	**98**	**92**	81	58	98	98	67	68	78	**98**	**98**	67	73	82
AB 11	100	68	68	72	53	68	67	**97**	**97**	68	68	68	70	78	68
AB 12		100	**92**	81	58	**98**	**98**	67	68	78	**99**	**99**	67	73	83
AB 14			100	82	57	**93**	**94**	67	68	80	**92**	**92**	66	72	82
AB 15				100	58	82	81	71	72	86	81	81	68	72	86
AB 16					100	58	58	52	52	57	58	58	52	54	60
AB 17						100	98	67	68	78	**99**	**98**	67	73	83
AB 19							100	66	67	78	**98**	**98**	68	73	83
AB 21								100	**96**	68	67	67	70	76	68
PuK 25									100	69	68	68	70	78	69
C. symbiosum										100	78	79	66	69	81
FFSB1											100	**99**	68	73	83
FRD38												100	68	74	83
H. salinarum													100	75	67
T. celer														100	71
TREC16															100